A Semiotic Systems Approach to Distributed Knowledge Environments

Cliff Joslyn

Computer Research and Applications Group (CIC-3)

Los Alamos National Laboratory

SRI International

- **Background:** Systems Science and Semiotics
- **Principia Cybernetica:** Distributed, collaborative self-application of cybernetic principles
- **LANL:**
  - –Distributed Knowledge Systems (DKS)
  - –Current research efforts
- Towards robust distributed knowledge environments

Systems Science

- Many meanings to many people:
  - An interdisciplinary approach to engineering and science from the perspective of "information systems"
  - Transdisciplinary study of the abstract organization of phenomena, independent of their substance, type, or scale
  - Search for isomorphisms among systems of different types
  - General theory of modeling
  - General science of complex systems
- History
  - War-time Birth: Early digital computers, feedback control
  - Post-war Growth: Model isomorphisms, computer technology
  - Backlash: Unfulfilled grand promises, postmodern philosophy, cooption
  - Modern Reflections: Complexity science, ALife

Reflections of Systems Science

- **Also known as:** Cybernetics, General Systems Theory
  - Weiner, Ashby, von Bertallanfy, von Foerster, Boulding, Beer, Klir, Pask, etc.
- **General precursor to:**
  - Systems engineering and analysis
  - Information theory, operations research
  - Connectionism
  - Dynamic systems (e.g. Club of Rome)
  - Complexity science (e.g. SFI)
  - Artificial life, theoretical biology
- **First transdisciplinary science:**
  - Work in breadth *and* depth
  - Few bases for consensus (texts)
  - Encyclopedic

Formal Approaches to General Systems

- **Dimensions:** The things that can vary; variables indeces; X_i
- **State Sets:** *How* these thing can vary (``data types"); $X_i = \{ x_{ij} \}$
- **System:** A multidimensional relation: $S \; f \; \mathrm{H}_i \, X_i$
- **Properties:** Of dimensions and relations:
  - **Cardinalities:** Finite, countable, uncountable
  - **Orderings:** None, partial, ordinal, numerical
  - **Boundedness:** Bounded, unbounded
  - **Relations:** Reflexive, symmetric, transitive, etc.
- **Projections:** consider only some dimensions at a time
- Connection to relational databases
- ***Most general formulation:***
  - Tradeoff of generality and expressibility
  - Isomorphic representations: graphs, equations, state machines

An Example

- Time: $X_1 = R$
  - Cardinal, unbounded, total order
  - Timeline
- Subject: X_2 = subject hierarchy
  - Nominal, unbounded, partial order
  - Node in tree
- Degree of association: $X_3 = [0,1]$
  - Cardinal, bounded, total order
  - Weight on link, slider position
- Record: triple in $X_1 \times X_2 \times X_3$
- Projections, examples:
  - $X_1, X_{2:}$ ignore weights
  - $X_1, X_{3:}$ shift in interest over time

Classical Information Systems Theory

- **Information** as the hypothetical universal concept for scientific unification
- **Related Concepts:** Order, organization, complexity, structure, form, hierarchy, relations, distinctions, variety, constraint, adaptation, evolution, development
- **Classical Measures:**
  - **Thermodynamic Entropy:** Measure of the distance of a physical system from thermodynamic equilibrium
  - **Statistical Entropy:** Measure of the uninformativeness, specificity, or ``spread'' of a probability distribution: $H(p) = - 3_i \, p_i \, log_2(p_i)$
- **Manifestations:** 2nd Law of Thermodynamics, 10th Theorem (Shannon), Law of Requisite Variety (Ashby), Maximum Entropy Principle (Jaynes)
- **Applications:** Coding theory, dynamical systems, dissipative structures, biological evolution and organization, genetic structures

Limitations of Classical Information Theory

- **Evident Limitations:**
  - Frame-dependence, measurability, conservation, type specificity, probabilistic basis
- **Greatest Concerns:** Information vs. Communication
  - *Real* information *means* something, is *used* to make decisions
  - Structural measures, merely quantitative
  - Shannon: ``These semantic aspects of communication are irrelevant to the engineering problem''.
  - The need for the elusive **semantic information theory** recognized for a long time (Bar-Hillel and Carnap, Minsky, Dretske, Harnad)
  - Denotational semantics (meta-syntactic labeling) necessary but not sufficient
- *What are the differences among data, information, and knowledge?*

Semiotics

- General theory of **representations:** *signs and symbols*
- **Origins:** Linguistics and humanities (philosophy of language)
  - Text and media analysis
  - Animal call systems
  - Theoretical biology
- **Concerns:**
  - Sign typologies
  - Digital/analog, symbolic/iconic representations
  - *Motivation:* intrinsic relations of sign to meaning
  - Mappings among representational systems, analogy, metaphor, category theory
- **Results:**
  - *Tokens:* creation, transmission, storage, *interpretation*
  - Modeling epistemology
  - Emphasis on *sources* of codes

Semiotic “Dimensions”

- **Syntax:**
  - Relations among *tokens*, production of new tokens
  - Usually formal
- **Semantics:**
  - Tokens (signs) *interpreted* by agent as standing-for environment observables
  - Measurement, actions
- **Pragmatics:**
  - Repercussions of sign interpretations for the agent in the environment
  - Purpose: goals, desires
  - Ultimate criteria: survival

Purposes of Principia Cybernetica

- **Primary:**
  - To develop the theory of Systems Science in general and MST Theory in particular: long-term, distributed, collaborative
  - Linkage of form and content: self-application of cybernetic technology to the development of cybernetic principles (echoing Whitehead and Russell)
  - To approach the conceptual unification of Systems Science by construction
- **Secondary:**
  - To develop methods to build consensus among communities of scholars
  - To flexibly represent material in many different forms
  - To balance editorial control with public participation to achieve both progress and openness
  - To move easily among natural language, formal language, and mathematical notation
  - To publish the whole or portions of the system through traditional means
  - To provide service to the Systems Science community

Ideal Environments: Universal Knowledge Systems

- Early visionaries: Vannevar Bush, Doug Engelbart, Ted Nelson
- Universal access
- Universal content
- Unlimited collaborative granularity
- Completely connectivity
- Completely flexible representations
- Maximal human interface
- Active, ``extensible'': merger of objects (data) and processes (programs) in a monadic meta-representation

What is Principia Cybernetica?

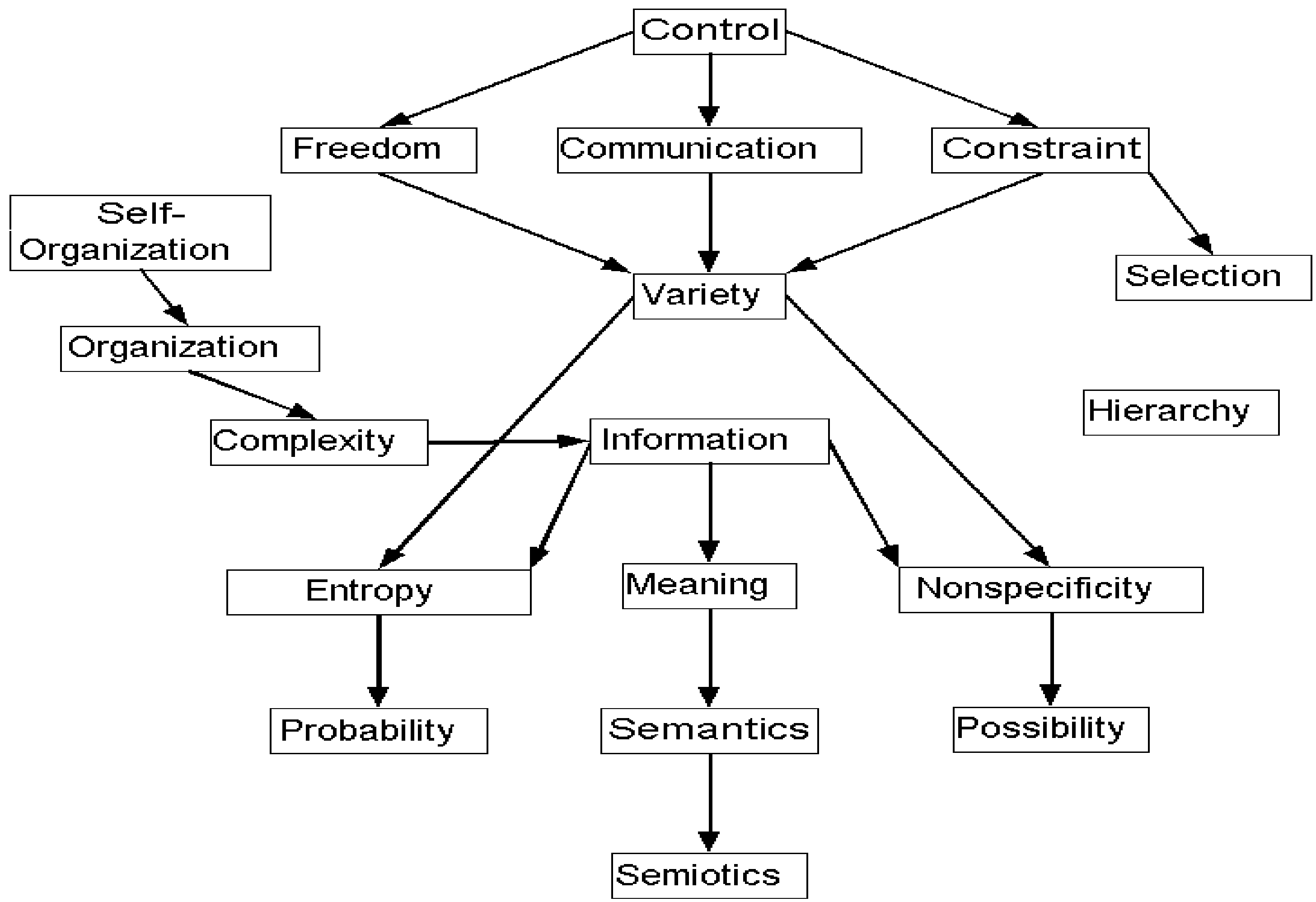
`http://pcp.vub.ac.be`

- A long-term, distributed, collaborative, international, scholarly project administered by an Editorial Board: Turchin, Heylighen, Joslyn
- A World Wide Web hypertext corpus for evolutionary cybernetics
- An experimental test-bed for ideas about evolutionary and semantic networks and webs
- An electronic home for Systems Science research and researchers
- An electronic publication allowing contributions by many member of an extended community

Principia Cybernetica Activities

- Philosophy and methods:
  - Reflexive linkage of form to content
  - Distributed, collaborative hypermedia *required*
  - Graph-theory based representations of semantic information
- 1989: Founding documents
- 1993: Web server: first in Belgium, clickable map, annotations, database generated corpus
- 1995: Early published work in adaptive hypertext
- 1996: Dictionary hosting, republishing (Ashby, Turchin)
- Current: Most authoritative web source for Systems Science
- Throughout:
  - Collaborative community
  - Students, publications, conference presence

Some Early Node Ideas

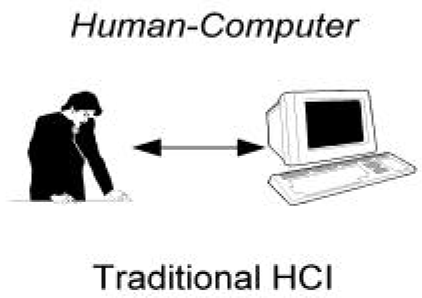


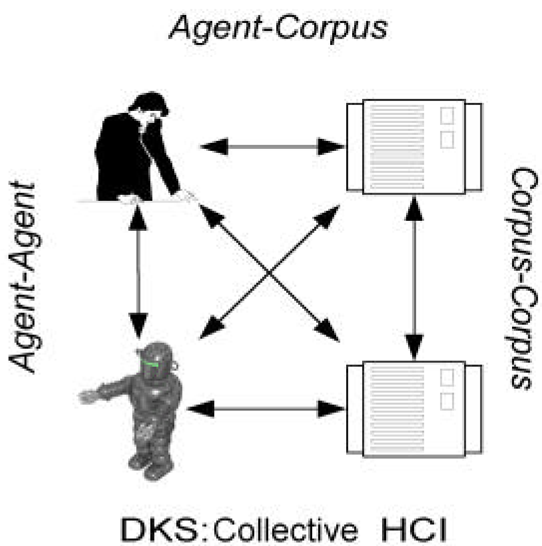
LANL Needs in Knowledge Systems

- **LANL priorities:**
  - Scientific and engineering knowledge bases:
    - Historical and current
    - Closed and open (Digital Library)
    - Creation, organization of, and retrieval from
    - In interaction with user communities
  - Bioinformatics: Genomics data, interacting experimental
  - Protection of the information infrastructure
  - *Scientific* approach to understanding these unprecedented new systems
- **National and global priorities:**
  - Unprecedented properties: combination of computation, storage, and communication
  - Potential to revolutionize the way society is organized
  - Promise to be the transformative technology of the 21st century

A Semiotic Systems View of Distributed Knowledge Systems (DKS)

- Knowledge or information
  - Data (bits on disks). . .
  - Interpreted by an agents (human and/or computational). . .
  - To help make choices, decisions, serve purposes
- DKS Defined:
  - Communities of interpreting agents
  - Interacting with networked information resources
  - Human-computer interaction at the *collective* level





More on a Broad Vision of DKS

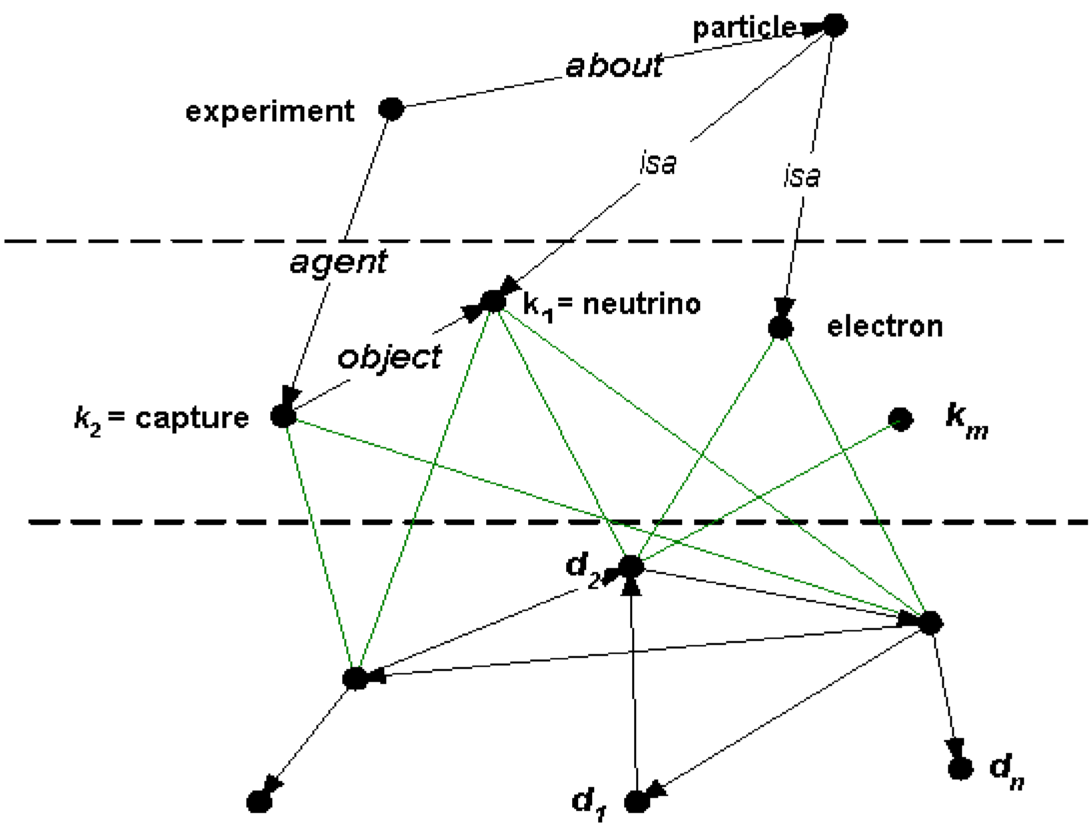
- Broad vision for DKS:
  - *Decentralized. . .*
  - Computer-assisted *information environments. . .*
  - For the representation and organization of *knowledge. . .*
  - Involving interactions with and among *communities. . .*
  - Of *users*, whether human, non-human, or autonomous computational processes
- Semiotic interaction in virtue of:
  - **Syntax:** Underlying computational processes
  - **Semantics:** The meaning of that information to user communities
  - **Pragmatics:** The use of that information by those communities to aid their stability, persistence, and survival

LANL Technical Orientation

- *Organization* for retrieval and extensibility
  - Beyond digitization
  - Explicit consideration of organization and structure
- *Interaction* with user communities (human and/or computational)
  - As source of information about structure
  - To aid in future retrieval
- Utilization of *tacit* or *implicit* information
  - Structural connections and hidden patterns
  - User action and interaction
- Representation of *uncertainty-based* information content
  - Subjective (assessed) and objective (measured) sources
  - Probabilistic and non-probabilistic mathematics
- Representation of *semantic* information
  - Information about context, meaning, ***interpretation***
  - At least keywords and meta-data, but beyond

Corpus Structure

Examples: Library Without Walls, Web



"Ontological":
Relations among keyterms
Authors
Communities of practice
Induced

Semantic:
Keyterms qualifying documents
Authors
Editors
Induced

Structural:
Links among documents
Citation
Linking
Traversal

LANL Research Efforts

`http://www.c3.lanl.gov/cic3/teams/knowledge`

- Natural Language Processing
  - NLP representations of semantic relations
  - Mapping to ontological categories and link types
- Recommendation Systems
  - `http://www.c3.lanl.gov/~rocha/lww`
  - Adaptive systems, uncertainty management
- Ontology Elicitation within Communities of Practice
- Sequence Analysis
  - User traversals, citation chains
  - Word sequences for topic identification
- Network Analysis
- **Robust Knowledge Management Environments**

Towards Portable, Robust Knowledge Environments

- Recognized need for portable movement of Knowledge Managment resources from one environment to another
- Yet KM systems *should* and *must* be specific to particular organizations
- KM Environments:
  - Instantiate in each organization
  - Develop within that organization
  - Different mix of tools and capabilities in each organization
- Purposes:
  - **Corpus Management and Representation:** Visualization, manipulation and analysis
  - **Natural Language:** Representation of semantic relations from text
  - **Ontological Representation:** Facilities for Communities Of Practice (COPs) to construct representations of their specific ontological structures.
  - **Data Mining:** Clustering, sequence analysis, network analysis, and other statistical tools to uncover hidden patterns and structure relations.
  - **Hypertext:** Authoring and analysis, semantically enhanced

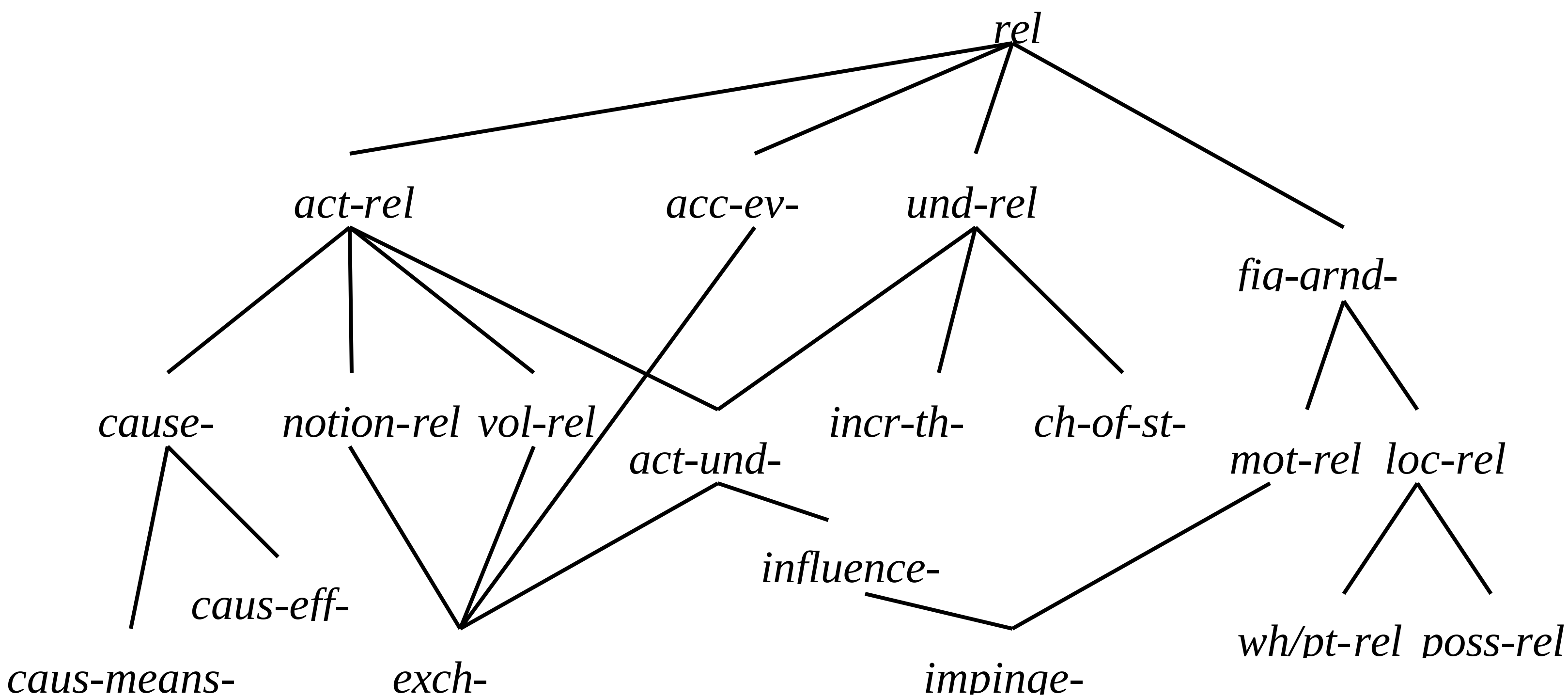
A Knowledge Representation Tool

- **Necessary components:**
  - Mathematical systems theory
  - Semantic information and relations
- **Graph theoretical basis:**
  - Labeled, weighted, directed graphs
  - Unlabeled, unweighted, simple graphs supported by default
  - Possible: category theory, hypergraphs
- **Interpretation:**
  - Nodes as projected states
    - Concepts, data, documents
  - Links as semantic relations: labels as link (relation) types
    - Categories: Classical (isa, hasa); discursive (support, refute); etc.: arbitrary
  - Weights: strength of relation

Link Type Architecture

- **Recovery of traditional concepts:**
  - Semantic networks
  - Weighted conceptual graphs (Sowa)
  - Entailment meshes (Pask)
- **Data typing mechanism:**
  - Relational properties:
    - Reflexivity, transitivity, symmetry
    - Their "antis“
    - Cycles
  - Identification: "how transitive is this sub-graph?“
  - Construction of closures: "save the transitive closure of this graph"
- **Multiple inheritance hierarchy:**
  - Inheritance of properties (e.g. transitivity), *and interpretations*
  - Semantic categories, predicates relating arguments (their nodes)
  - Real ontological or semantic relations

Example: NLP Semantic Relations



Davis 1995

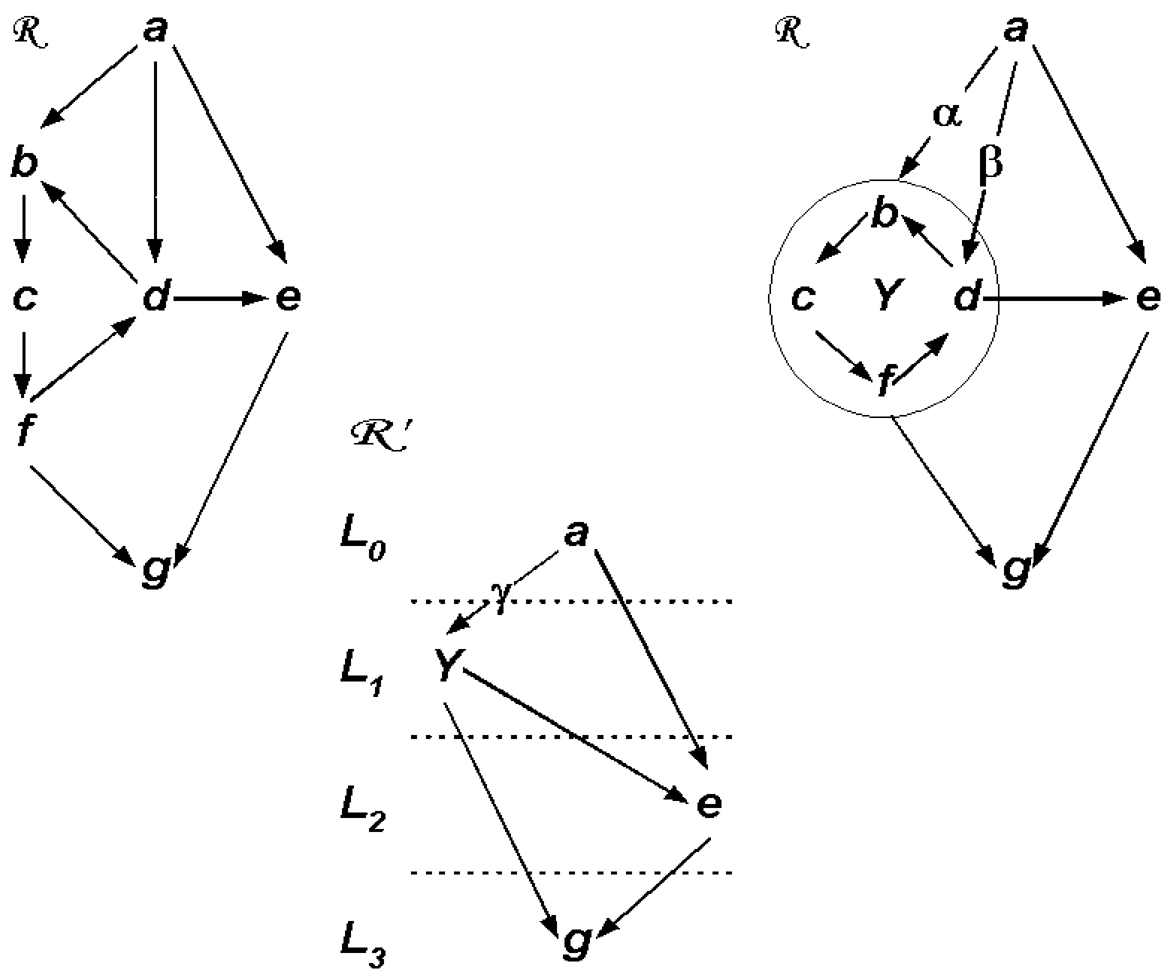
Software Engineering Environment

- **Front End:**
  - Full drawing/visualization environment for node-arc diagrams
  - Rubberbanding: Visio
  - Moden software interface standards: a robust GUI,
  - Full audit trail (time stamping, authoring, versioning)
  - Graphical and textual annotation, etc.
- **Back Ends:**
  - **Markup Language:** Full read/write compatibility
  - **Database:** Relational/OO architecture to record each graph
  - **Specialized Graph-Theoretical:** For transfer to e.g. Matlab

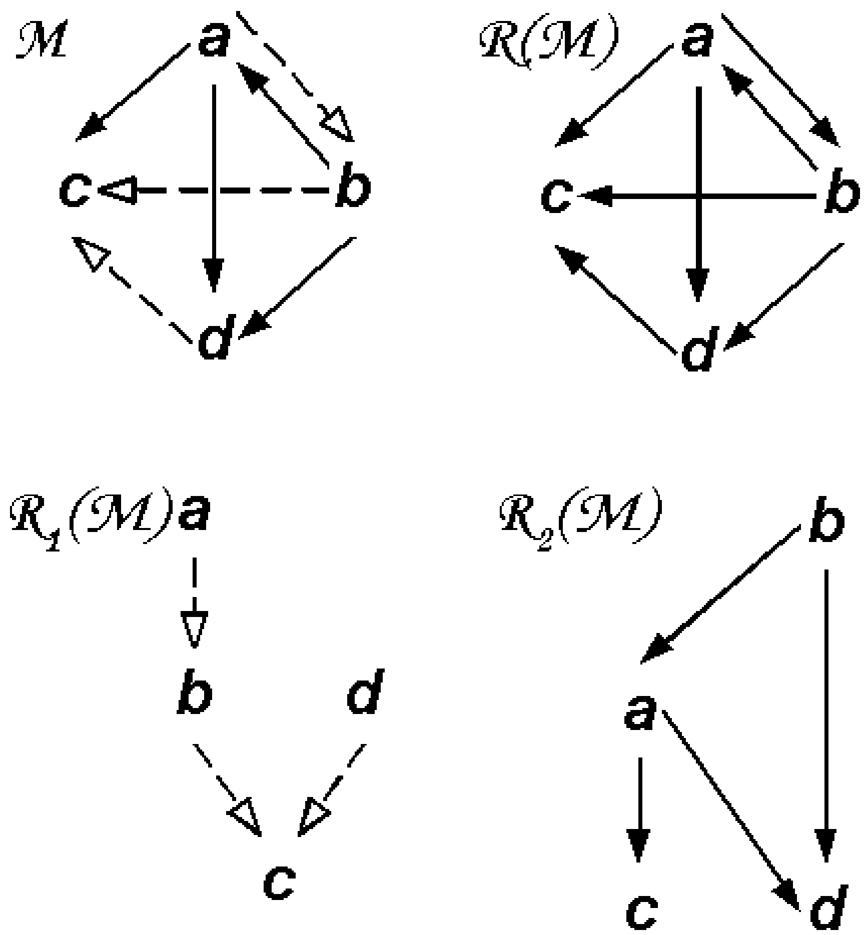
Graph Theory Support

- Subgraph Extraction: by link type, node type, n-neighborhood, or a cut-level of weight.
- Dual graph construction
- Principle component analysis
- Cycle-finding ("Is this graph cyclic? Show me the cycles. Reduce the cycles to new meta-nodes")
- Chain-finding ("Show the linear chains from A to B".)
- Shortest path General and standard graph statistics
- Metricity calculations
- Planarity optimization
- Clustering
- Root and leaf finding
- Level-finding (when the graph is DAG)
- Morphisms, equality testing, distance measures ("how similar are these graphs? can I twist this one into that one?")

Example: Cyclic Reduction



Example: Multirelations



Conclusions: Questions on Moving Forward

- Connect to current Knowledge Representation work:
  - Development environments for Conceptual Graphs?
  - Better mathematical development: graphs, relational theory, inheritance
- Connect more strongly to current markup standards:
  - DOM, XML, XSchema
  - KQML, KIF
- Development questions:
  - Architecture
  - Tractibility?
  - Development strategy
- LANL support